Probabilistic Index Models

Jan De Neve[1], Olivier Thas[1], Lieven Clement[2] and Jean-Pierre Ottoy[1]

[1] Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Belgium
[2] L-BioStat, K.U. Leuven, Belgium

Abstract

Recently the probabilistic index (PI), which is defined as the probability $P\{Y \leq Y'\}$ with Y and Y' independent random variables, acquired an increase of interest in Biostatistics. The PI is accepted as an informative and intuitive index for quantifying treatment effects, particularly when the treatment does not only act on the mean response. In a clinical trial for example, the effect of increased dosage of a drug on a response can affect the mean, dispersion and even the tails of the density. The PI summarizes the effect on the shape of the distribution into a simple effect measure. See for example, [1, 2, 3]. Most of these methods however, rely on distributional assumptions and they are limited to comparing two groups while correcting for additional covariates. We propose a much more general model for the PI,

$$P\{Y \leq Y'|X, X'\} = m(X, X'; \beta),$$

where (Y, X) and (Y', X') are independent and identically distributed random vectors and m defines the relation between the probabilistic index and the predictors. This general representation makes the model applicable for a variety of designs. For example, if Y denotes the blood pressure, X the dosage of a drug and the model $P\{Y \leq Y'\} = \text{expit}\{\beta(X' - X)\}$ is considered, then $\text{expit}\{\beta\}$ gives the probability that the blood pressure increases when the dosage is increased by one unit. Hence the model allows for a meaningful effect quantification of a continuous predictor on the response distribution, without making any restrictive distributional assumptions. The model can easily be extended to multiple predictors. It is shown that the framework naturally give rise to some traditional rank tests such as the Wilcoxon-Rank-Sum, Kruskal-Wallis and Friedman-Rank test and it allows extending these tests to more complicated designs. Asymptotic normality of the estimators and consistency of the covariance matrix estimator are established through semiparametric theory. The model is extensively illustrated on examples and the estimation theory is validated in a simulation study.

References

[1] Brumback, C.L. Pepe, M.S. and Alonzo, T.A. (2006). Using the ROC curve for gauging treatment effect in clinical trials. *Statistics in medicine* 25, 575–590.

[2] Dodd, L.E. and Pepe, M.S. (2003). Semiparametric Regression for the Area Under the Receiver Operating Characteristic Curve. *Journal of the American Statistical Association* 98(462), 409–417.

[3] Tian, L. (2008). Confidence intervals for $P(Y_1 > Y_2)$ with normal outcomes in linear models. *Statistics in medicine*, 27 4221–4237.